Predicting Nonspecific Ion Binding Using DelPhi

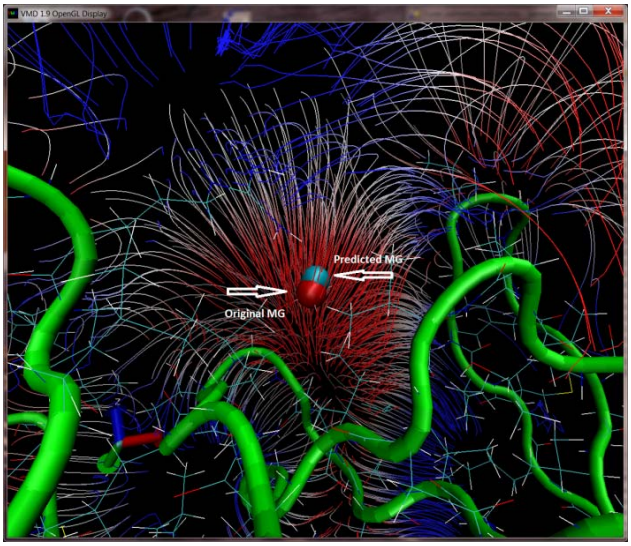
Marharyta Petukh, Maxim Zhenirovskyy, Chuan Li, Lin Li, Lin Wang, and Emil Alexov

Computational Biophysics and Bioinformatics, Department of Physics and Astronomy, Clemson University, Clemson, South Carolina

SUPPORTING MATERIALS

1) Illustration of narrow potential valley (*a*) and shallow potential valley (*b*).

(a)



(b)

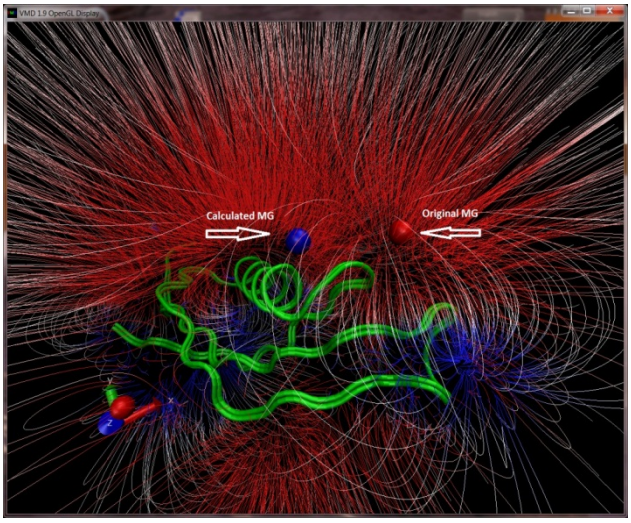


Figure 1S (*a*) – illustration of sharp potential well with experimentally observed ion in the potential minimum (protein ID 2hb4) and (*b*) illustration of shallow potential valley with experimentally observer ion at the near minimum potential value (protein ID 2cc9).

Visualization was made with VMD software (1). Electrostatic field of protein is represented as "ForceField" lines, where *red* color represents area with negative potential and *blue* – with positive one.

2) Analyzing the accuracy of the method in order to detect the experimentally determined position of tested ions

From the *Rank* distribution of cluster representative grid point with shortest distance to experimentally determined ion position (*Dmin, light bars*) it can be seen that in the vast majority of the cases, the representative grid point is located within 10 Å from the actual ion's position, that is equal to the uncertainty of clustering method. This assures that the clustering algorithm does not eliminate potentially good representative candidate points. The maxima of the distributions of the distance between the ion's actual position and the representative grid point with highest absolute value of the potential (*D(Pmin/max), dark bars*) are within 30-40 Å for all types of ions.

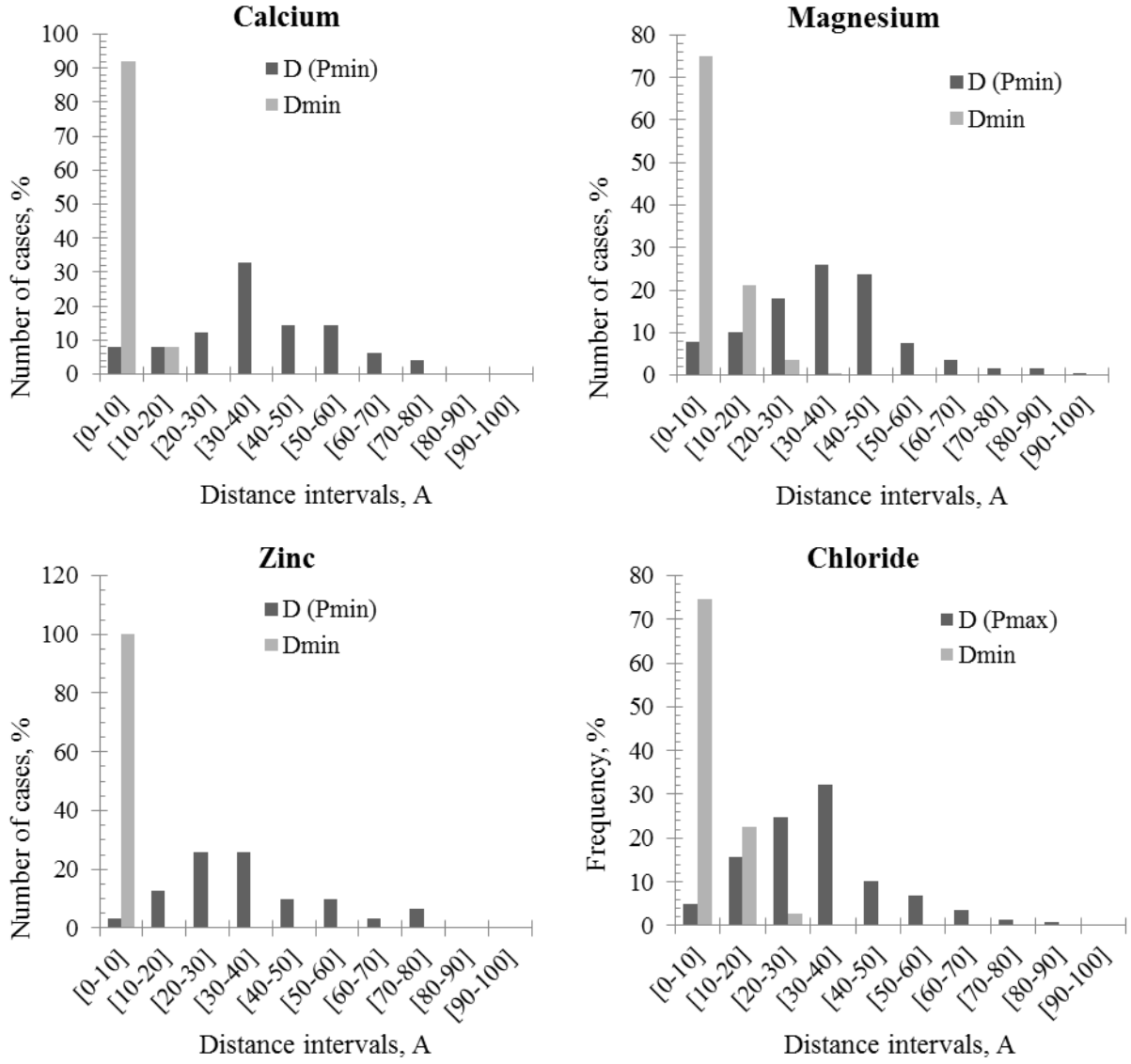
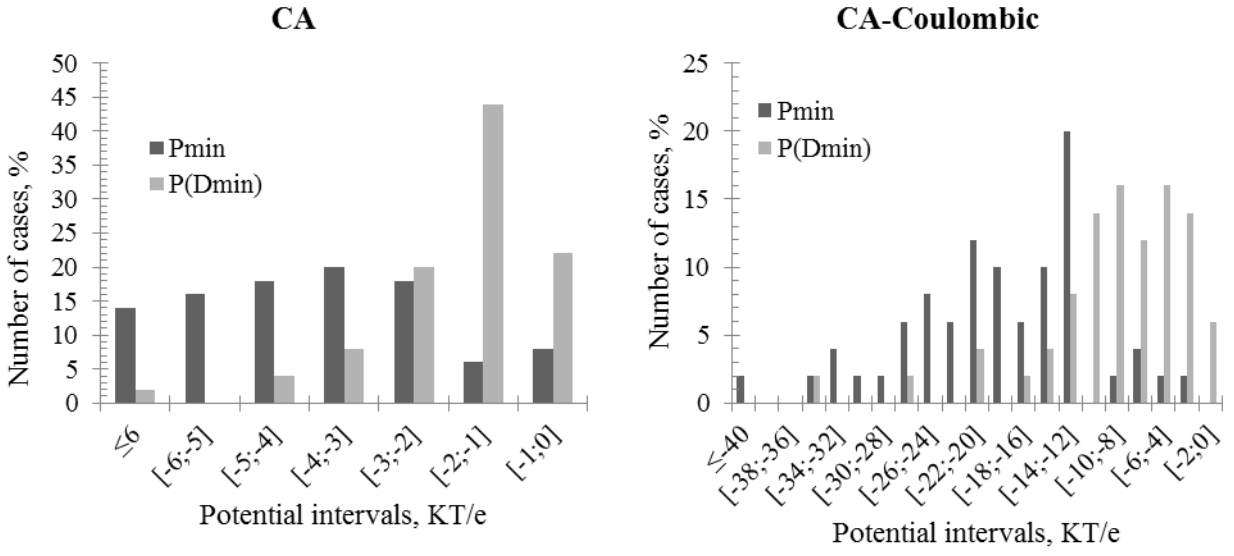


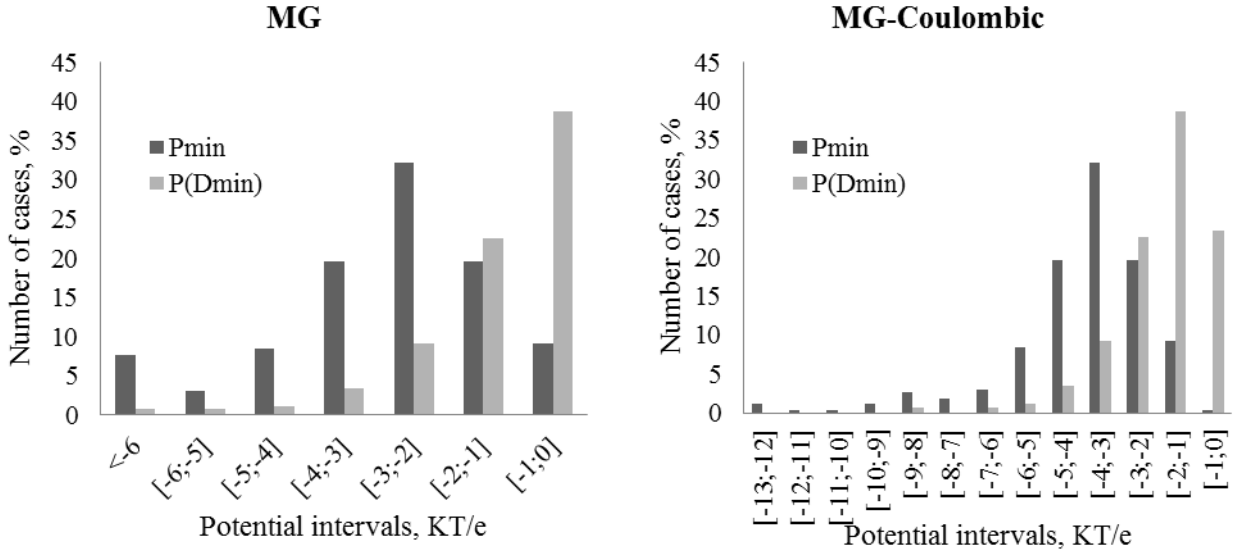
Figure 2S. Distribution of distances between the actual position of the ion and the representative grid point with highest absolute value of the potential (*Rank* = 1) (*dark bars*) and distances between the ion's original position and closest representative grid point (*light bars*).

3) Comparison of the predictions made with DelPhi (left panels) and with Coulombic law (right panels)

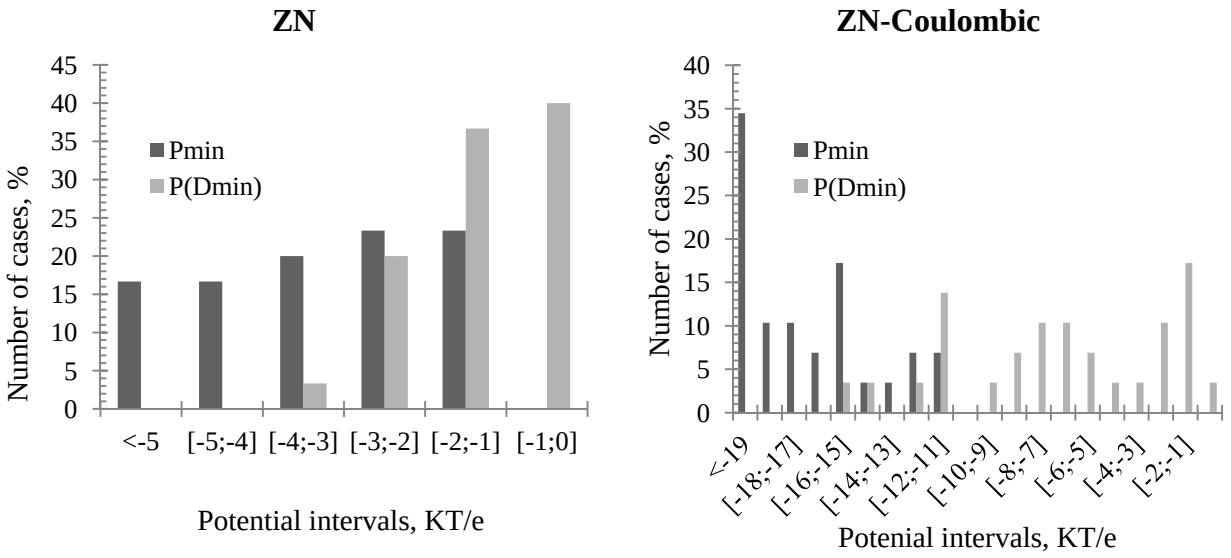
The graphs below show the distribution of two quantities discussed in the text: maximal (for Cl ions) and minimal (for other ions) potential value (*Pmin/max*) and the distribution of the potential at the point closest to the experimental position of the ion *P(Dmin)*. In the best case scenario, these distributions should be similar to assure good predictions. However, the graphs below indicate significant differences when the potentials are calculated via Coulombic law, especially for some types of ions, as for example the Zn and Cl ions. The corresponding distributions are drastically different suggesting that Coulombic law is not sufficient to generate correct predictions.



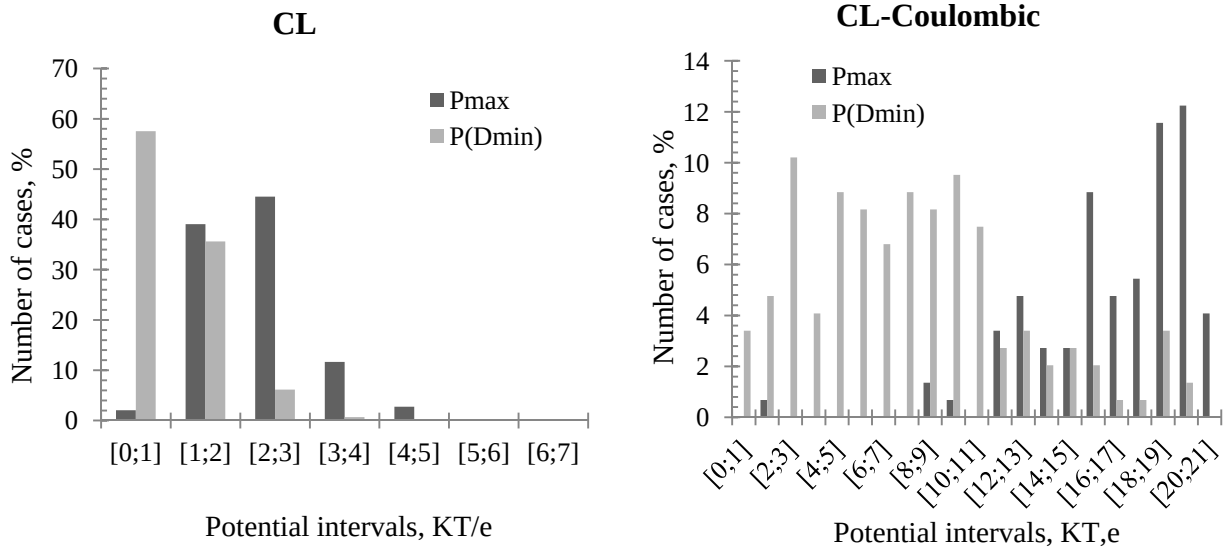
(a)



(b)



(c)



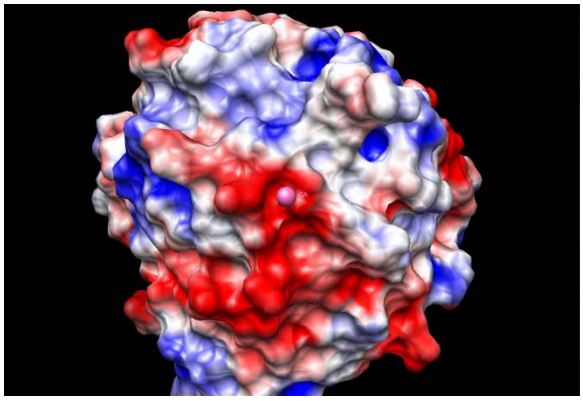
(d)

Figure 3S. Distribution of absolute maximum potential (*dark bars*, *Pmin for Ca*, *Mg and Zn; and Pmax for Cl*) and the potential of the representative grid point closest to the original corresponded ion position (*light bars, P(Dmin)*) for Ca (*a*), Mg (*b*), Zn (*c*) and Cl (*d*) containing experimental dataset. Calculations were made with different parameters. (*Left panel*): considered all type of molecular interactions, dielectric constant of solution 4, ionic strength 0.15 M; (*right panel*): considered only Coulombic interactions with dielectric constant 4 and ionic strength 0.15 M.

4) Analysis of prominent failures

In this paragraph we show worst predictions made for each type of ion and analyze the plausible reason for these failures.

(a)



(b)

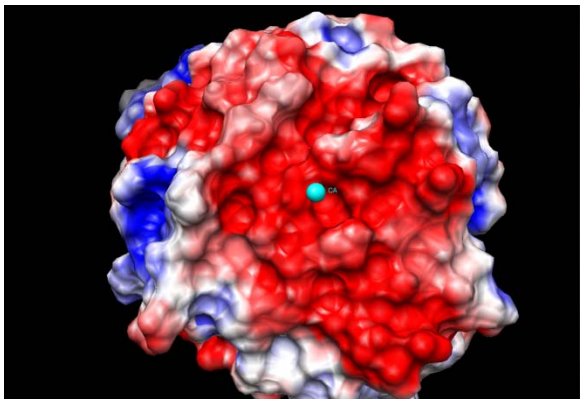
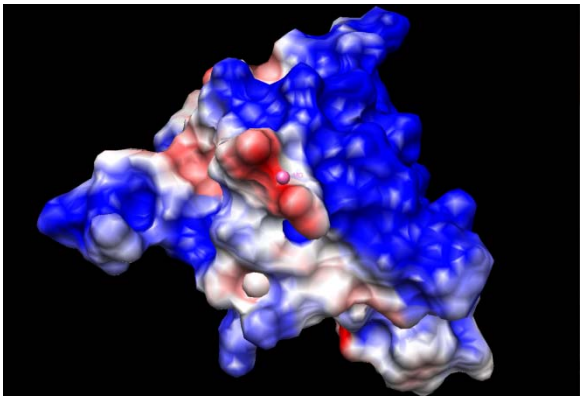


Figure 4S. Illustration of electrostatic potential distribution on the surface of *calcium*-containing protein (ID 1iz7) with experimentally determined ion (*pink*) (*a*); and ion, placed at the position of representative grid point with *Rank=1* (*cyan*) (*b*). Images (*a*) and (*b*) represent the front and the back of the protein. One can see that the predicted ion is placed inside the shallow valley with strong negative potential, while the experimentally determined ion is located inside narrow potential valley. In this case, adding geometrical constrains into the scoring function would to favor deep and narrow valley of the potential versus stronger potential but forming shallow valley would improve the predictions.

Visualization was made with Chimera software (2), where *red* color represents area with negative potential (min -2 KT/e) and *blue* – with positive one (max 2 KT/e).

(a)



(b)

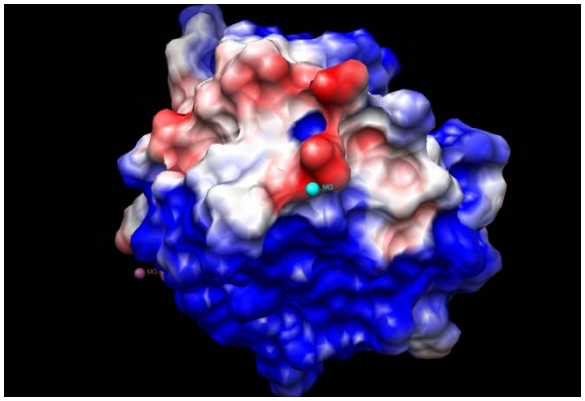
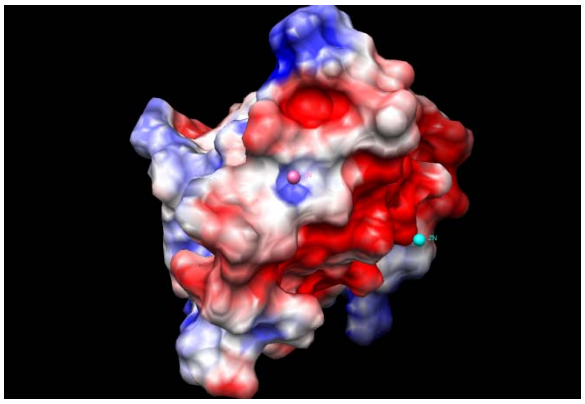


Figure 5S. Illustration of electrostatic potential distribution on the surface of *magnesium*-containing protein (ID 1l9a) with experimentally determined ion (*pink*) (*a*); and ion, placed at the position of representative grid point with *Rank=1* (*cyan*) (*b*). Images (*a*) and (*b*) represent different sites of the macromolecule. The electrostatic potential distribution on the surface of protein indicates that there are only few spots with negative potential. The experimentally determined ion is located in the surface cavity of one of these spots. The predicted position is located at stronger potential, but is not geometrically fitted. The experimental position is located in a cavity which can be speculated that additionally immobilized the ion. Thus, in this case, incorporating into the scoring function the molecular surface shape would improve the predictions.

Visualization was made with Chimera software (2), where *red* color represents area with negative potential (min -2 KT/e) and *blue* – with positive one (max 2 KT/e).

(a)



(b)

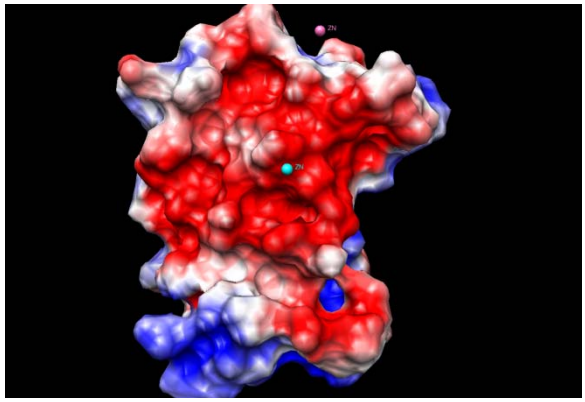
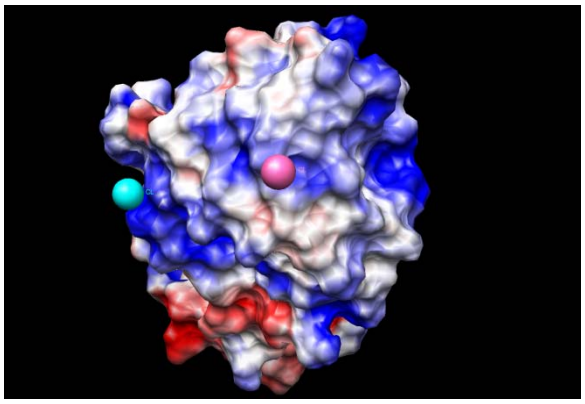


Figure 6S. Illustration of electrostatic potential distribution on the surface of *zinc*-containing protein (ID 3byr) with experimentally determined ion in the center (*pink*) and predicted ion on the side (*cyan*) (*a*); and ion, placed at the position of representative grid point with *Rank=1* in the center of image (*cyan*) and experimentally determined ion (*pink*) on the top (*b*). Images (*a*) and (*b*) represent different sites of the macromolecule. As shown on the image, the position of predicted ion is more preferable from the point of electrostatics and ion solvent accessible surface area. According to the experimental data the ion is located inside the cavity with weak negative potential around is and positive potential at the bottom of cavity. Thus electrostatics cannot predict this binding site and obviously other factors, not included in our protocol, are responsible for the binding.

Visualization was made with Chimera software (2), where *red* color represents area with negative potential (min -2 KT/e) and *blue* – with positive one (max 2 KT/e).

(a)



(b)

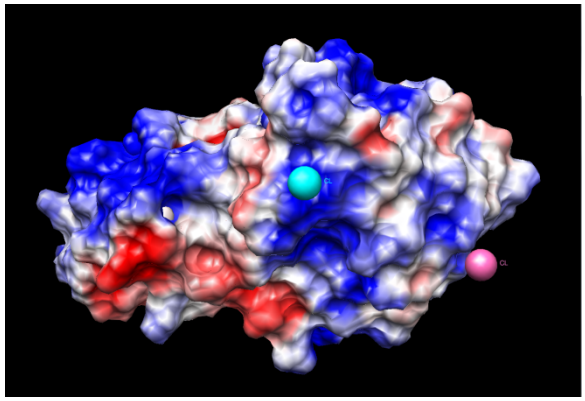


Figure 7S. Illustration of electrostatic potential distribution on the surface of *chloride*-containing protein (ID 160l) with experimentally determined ion in the center of image (*pink*) and predicted ion on the side (*cyan*) (*a*); and ion, placed at the position of representative grid point with *Rank=1* in the center of image (*cyan*) and experimentally determined ion (*pink*) on the side (*b*). Images (*a*) and (*b*) represent different sites of the macromolecule. One can notice that the predicted ion is placed at the shallow valley with strong positive potential, while the experimentally determined ion is located at the area with weaker potential. Obviously other contributions different from electrostatics are the driving forces for this particular binding.

Visualization was made with Chimera software (2), where *red* color represents area with negative potential (min -2 KT/e) and *blue* – with positive one (max 2 KT/e).

5) Receiver operating characteristic (ROC) curves

To test the sensitivity of the method to several parameters used in protocol (the internal dielectric constant of protein and the ionic strength in the water phase) and to find the optimal values of these parameters we constructed *ROC curve* (as described in Method section). The graphs below show ROC curves for each type of ion, received by choosing two different values for the internal dielectric constant 2 and 4 and ionic strength 0.15 M and 0.5 M.

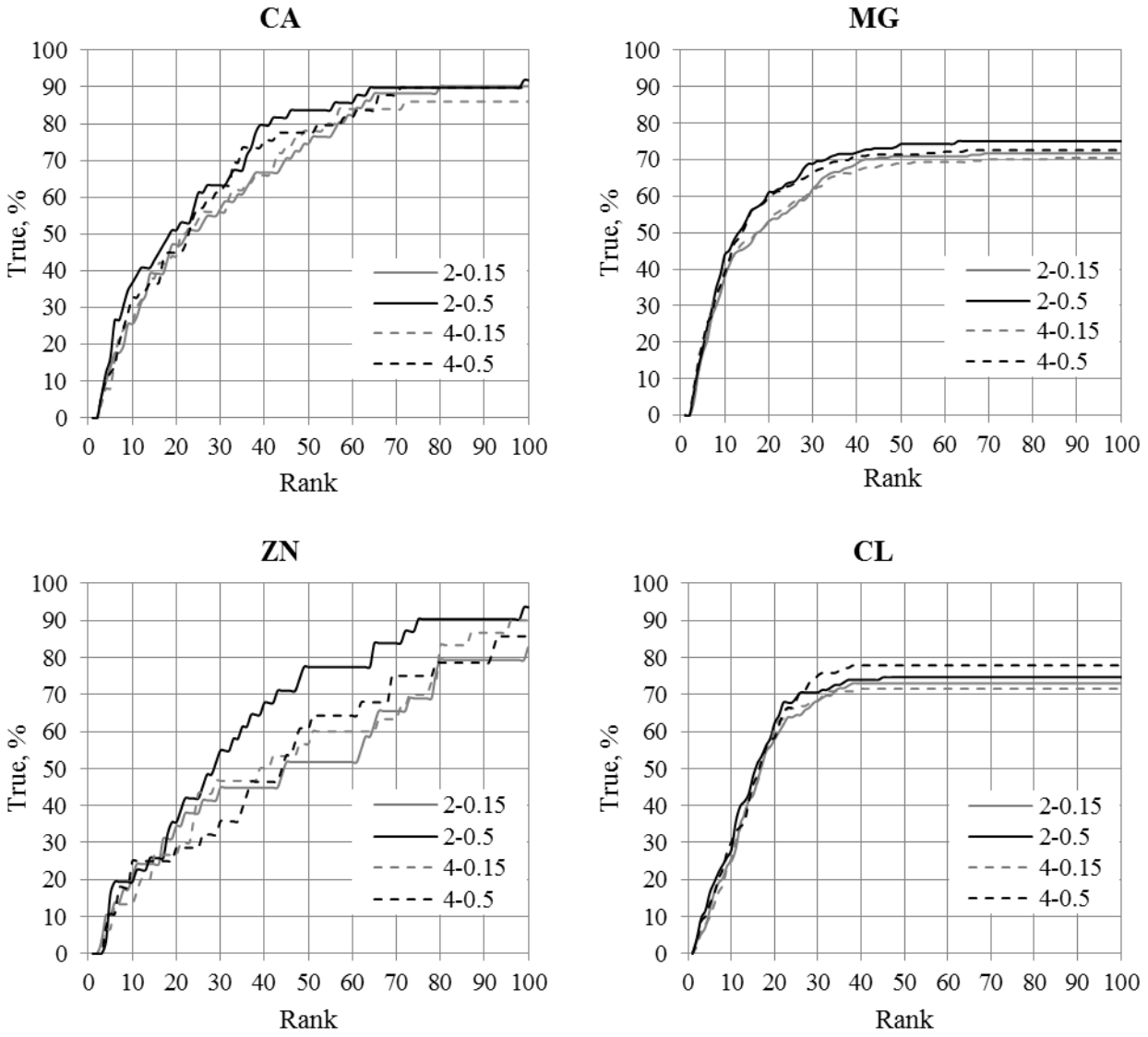


Figure 8S. ROC curves for Ca, Mg, Zn and Cl ions containing proteins dataset, calculated with respect to different parameters. In *legend* the first number corresponds to the dielectric constant of the solution; the second one – to the ionic strength in moles/l. The *x-axis* represents the *Rank* of the closest representative grid point to the experimentally determined ion position; *y-axis* – the number of true predictions in percentage of all predictions. A prediction is considered to be true if the distance between the predicted representative grid point and the actual experimental ion position is less than 10 Å. It can be seen that the method is not sensitive to the values of the parameters in cases of Ca, Cl and Mg ions, but quite sensitive in case of Zn ion. For all types of ions the best results are obtained with internal dielectric constant equal to 2 and ionic strength equal to 0.5M.

SUPPORTING REFERENCES.

1. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.
2. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-1612.